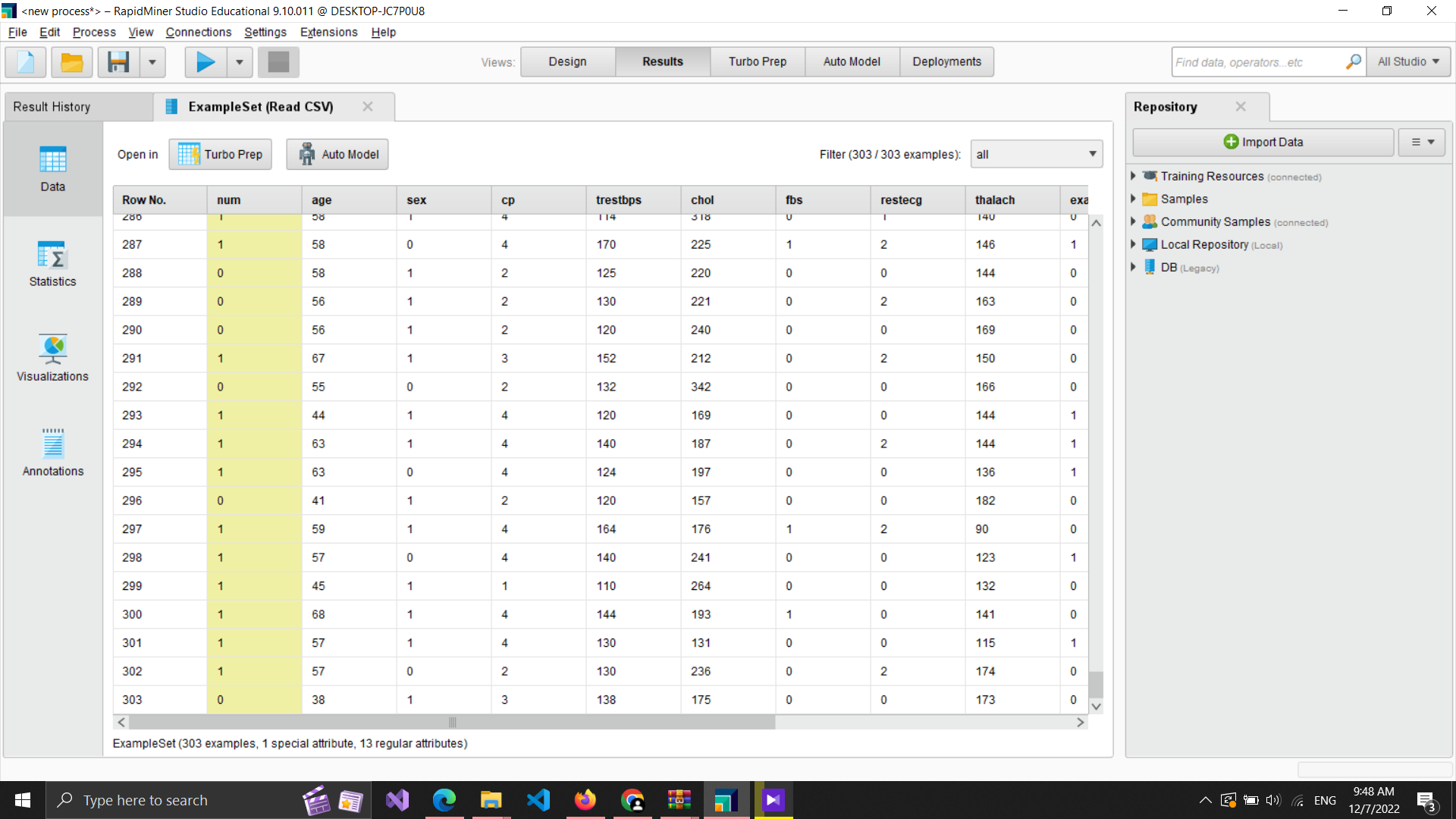
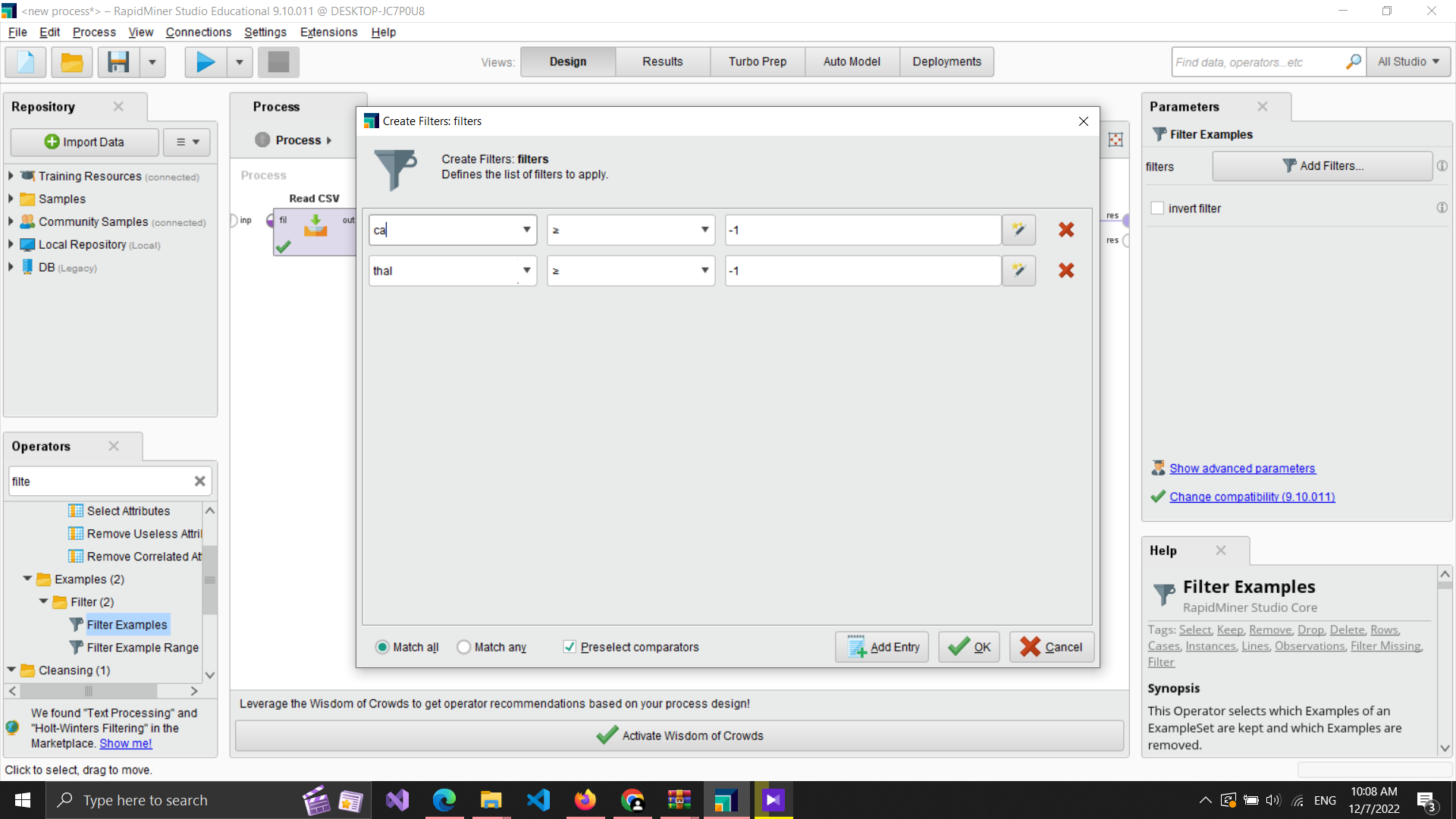
زهراحیدری 986203146

در این پروژه یک فایل از لیست بیماری های قلبی است که مواردی در آن مشخص شده

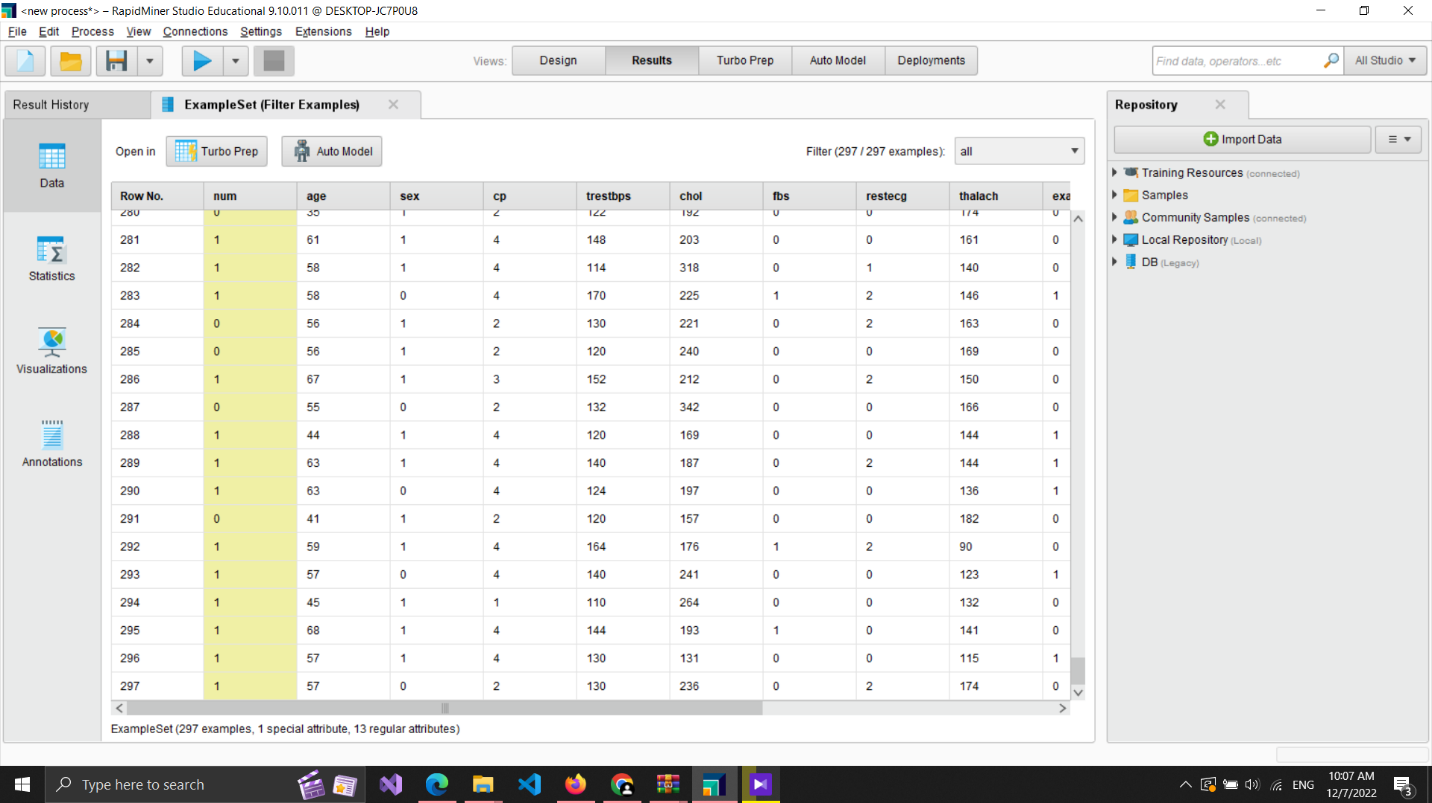
سن فرد  
 2- جنسیت (1 male و 2 female)  
 3- نوع درد در قفسه سینه  
 4- فشار خون  
 5- کلسترول  
 6- بیماری قند دارد یا نه  
 7- نتیجه دستگاه ضربه نگار قلب  
 8- حداکثر ضربان قلب  
 9- نتیجه تست ورزش  
 10 تا 13 سایر آزمایش ها  
 14- ستون برچسب کلاس که می تواند صفر به معنی بیماری قلبی و یک  بیماری قلبی ندارد.

که شامل 303 سطر میباشد که با استفاده از نرم افزار rapidminer ما فایل اکسل آن را میخوانیم

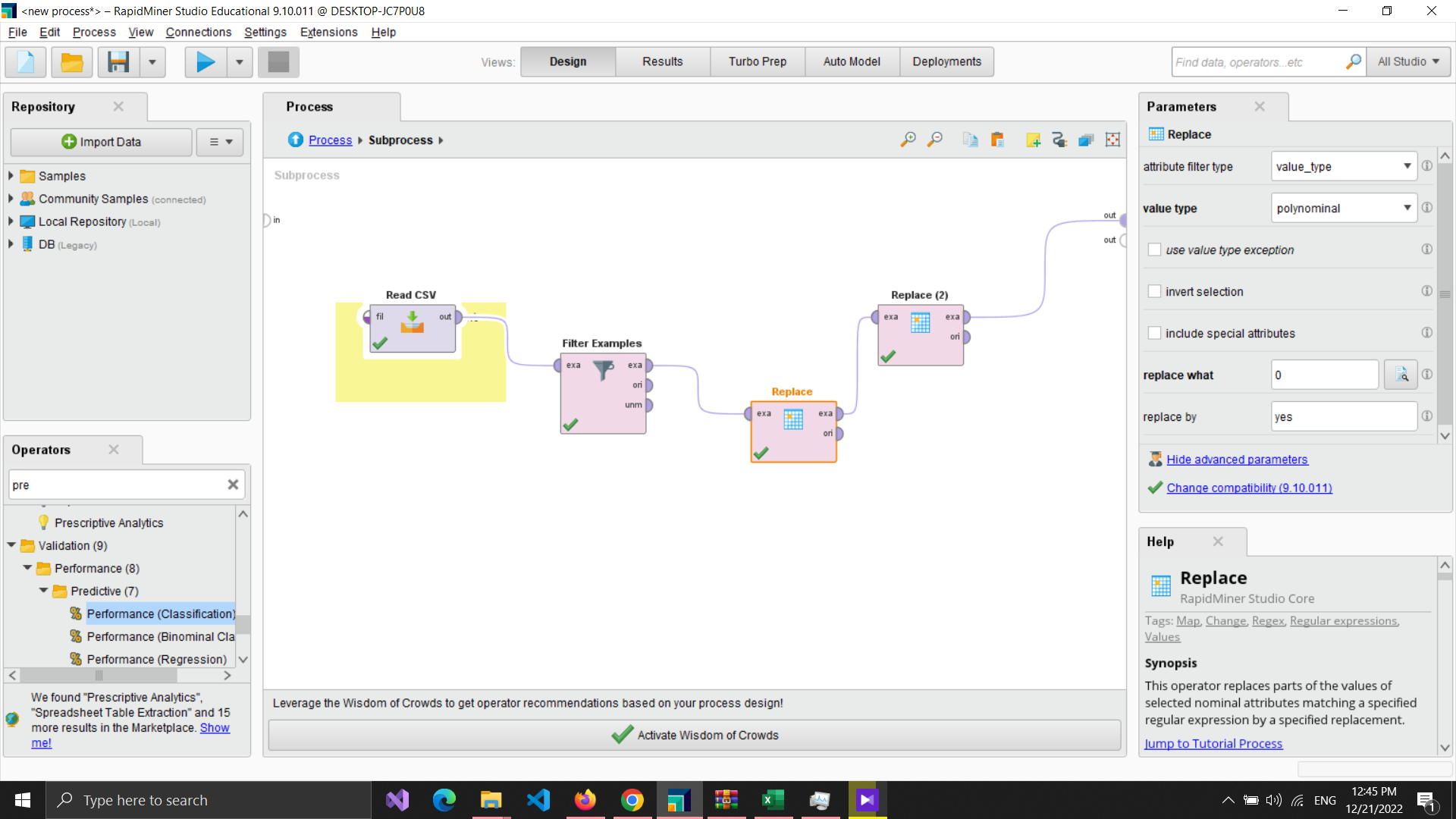


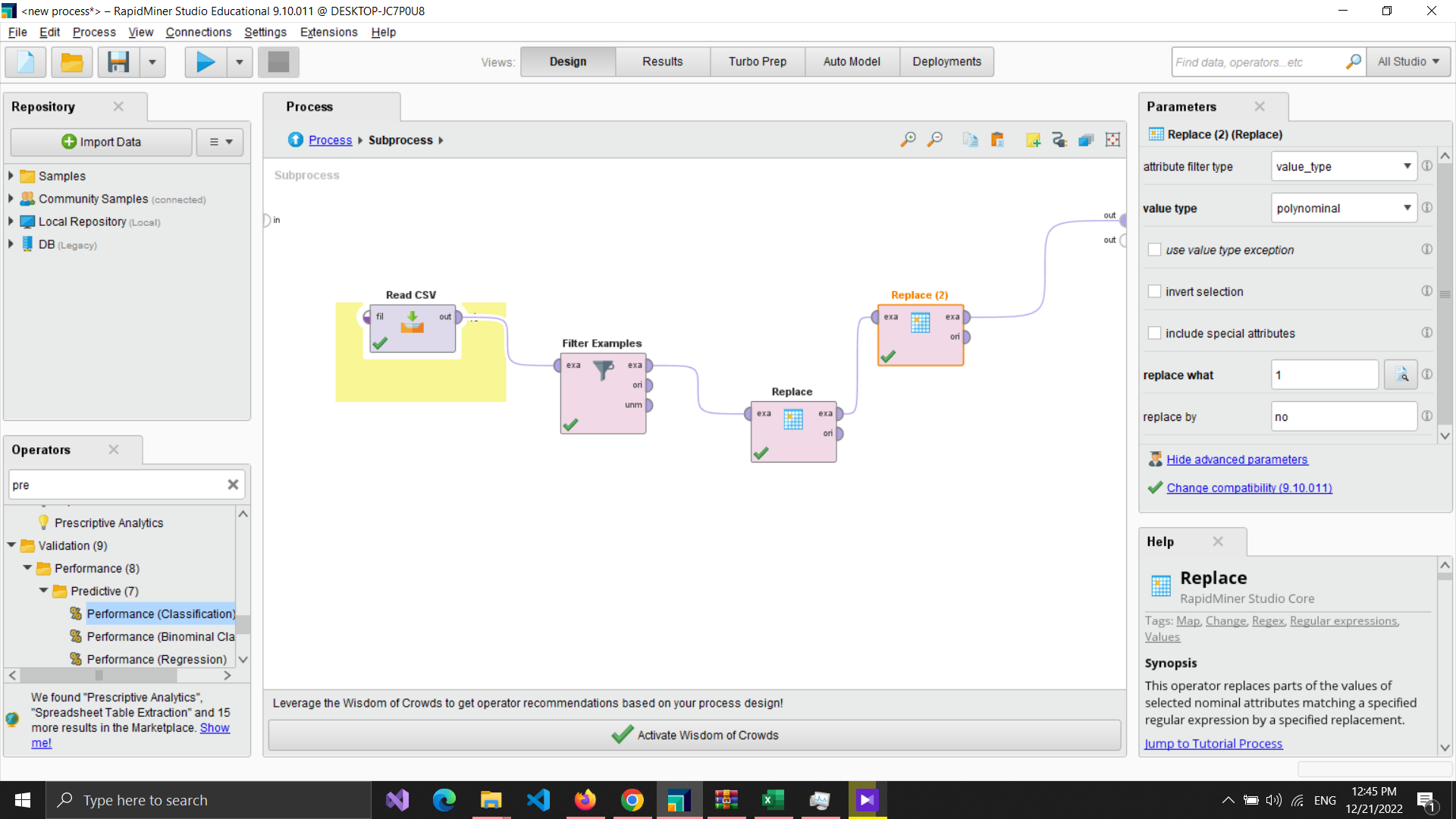
در ادامه با مشاهده کلی داده ها متوجه یه سری داده های اشتباه میشویم. که در ستون ca ,thal دیده میشود که عدد منفی 1000 را نمایش میدهد برای همین ما کلا با ابزار فیلتر می آییم ان هایی که منفی هستن از این دو سطر به طور کلی حذف میکنیم (یا حتی میشود با replace مقدار میانگین را برابر این مقدار قرار داد) 

که داده ها به 297 تا تغییر میکند یعنی 6 تا از داده ها فیلتر میشود.

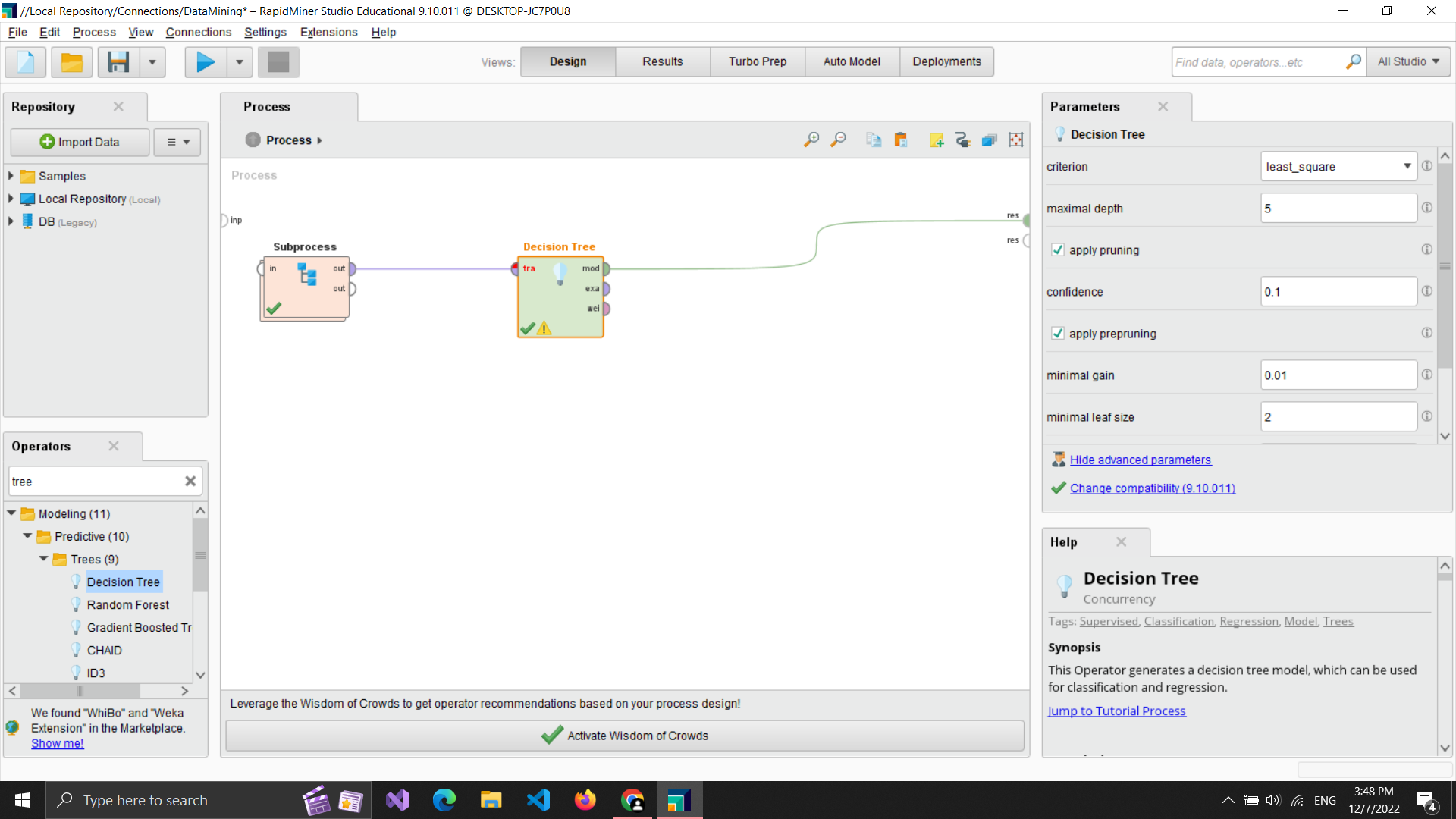


و سپس من یه تغییری در داده ها دادم اینک اون داده 14 که لیب میشود به اضافه 3تا ستون دیگر را بجای 0 و 1 به ترتیب مقدار yes و no قرار دادم که برای آسانی خودم برای داده هاست.

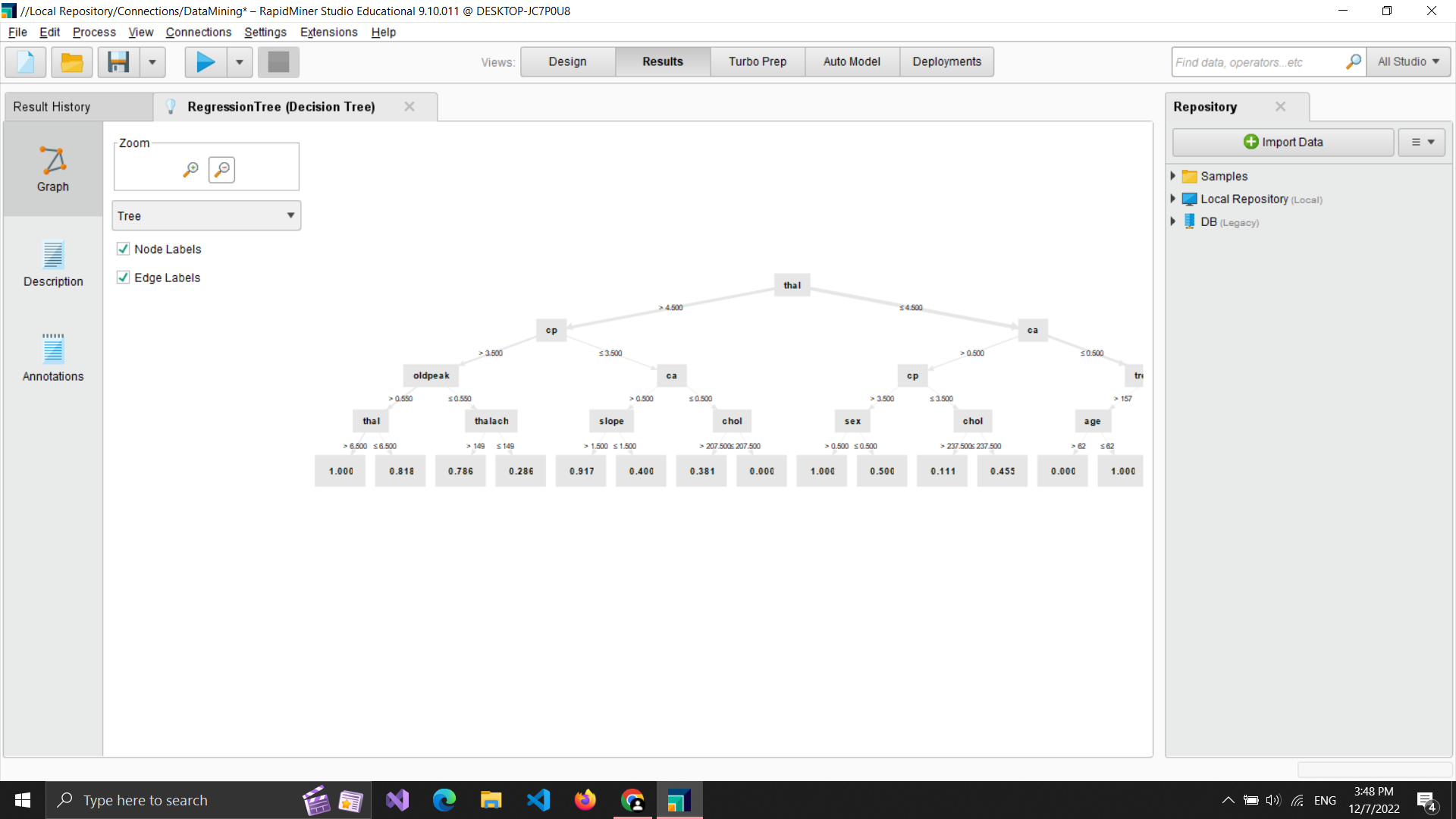


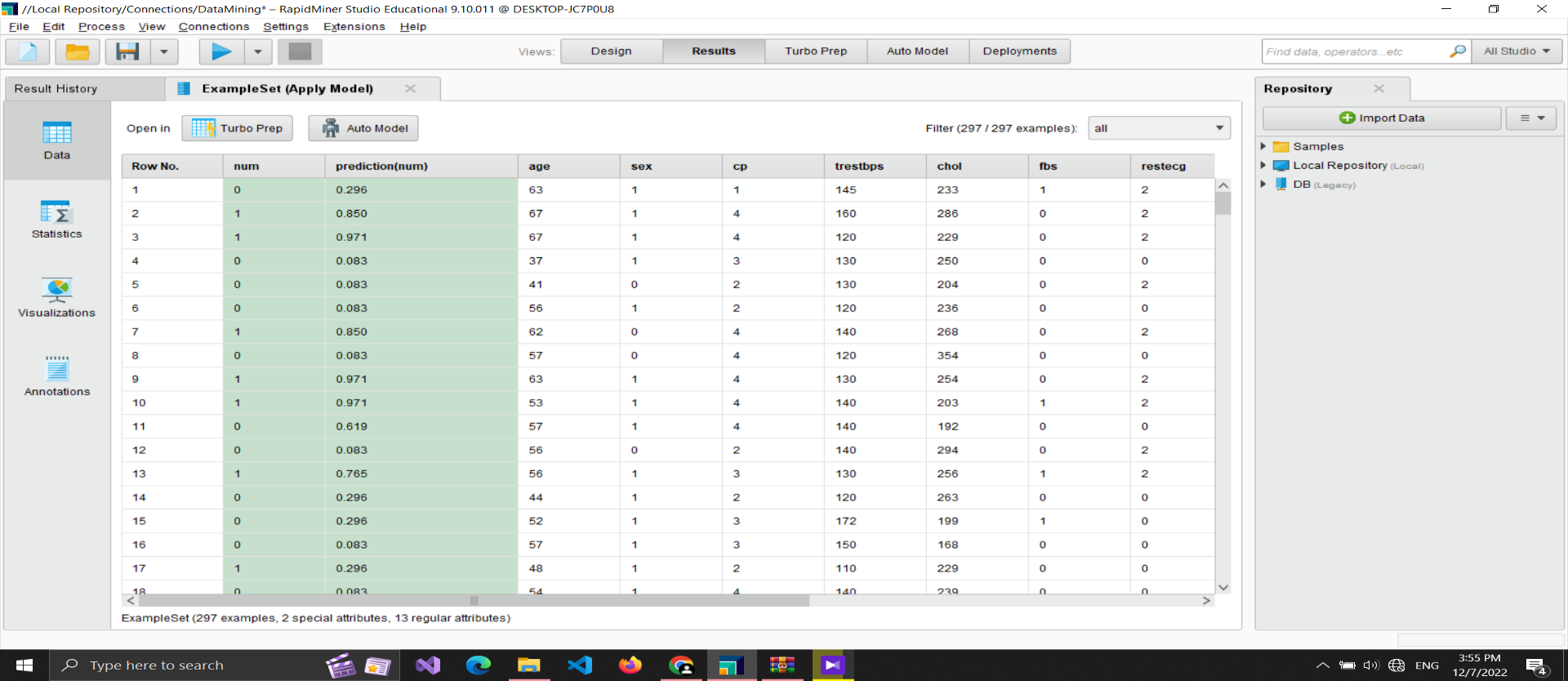


و همه این موارد را در subprocess قرار میدهیم. که یکم دیدن شمای برنامه راحت تر شود.

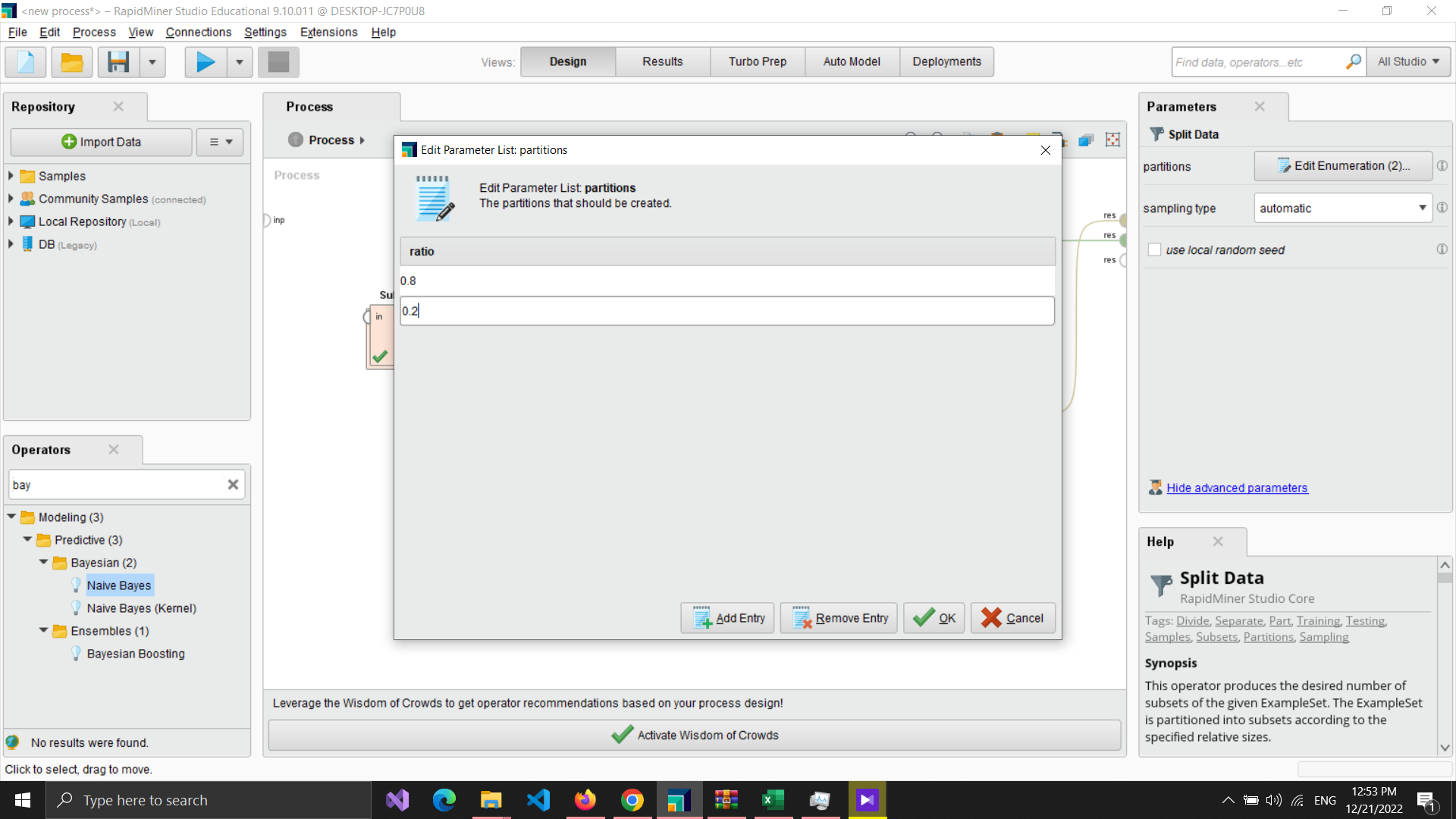


ابزار درخت تصمیم را به آن اضافه میکنیم . و خروجی آن به صورت زیر است.

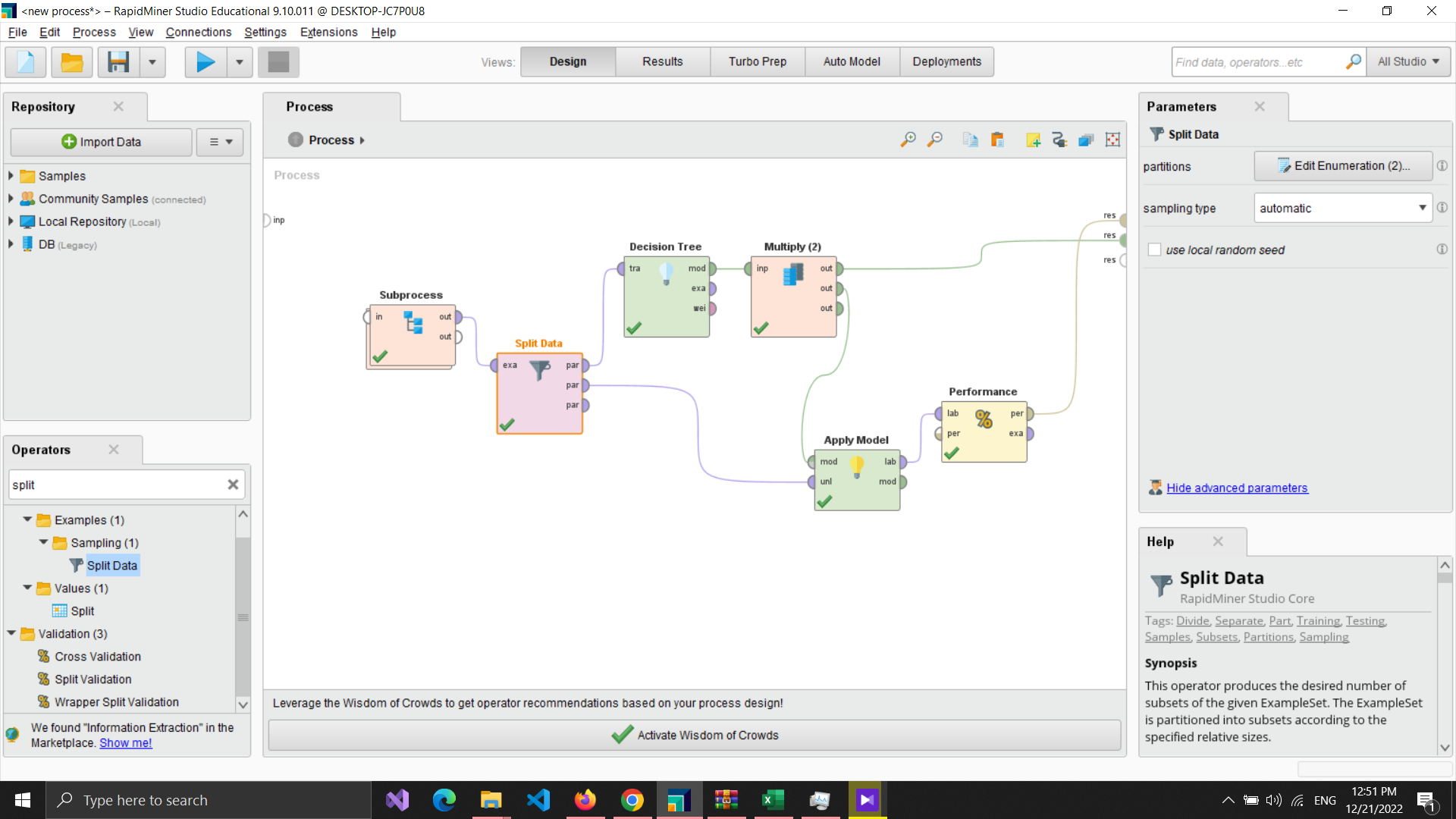




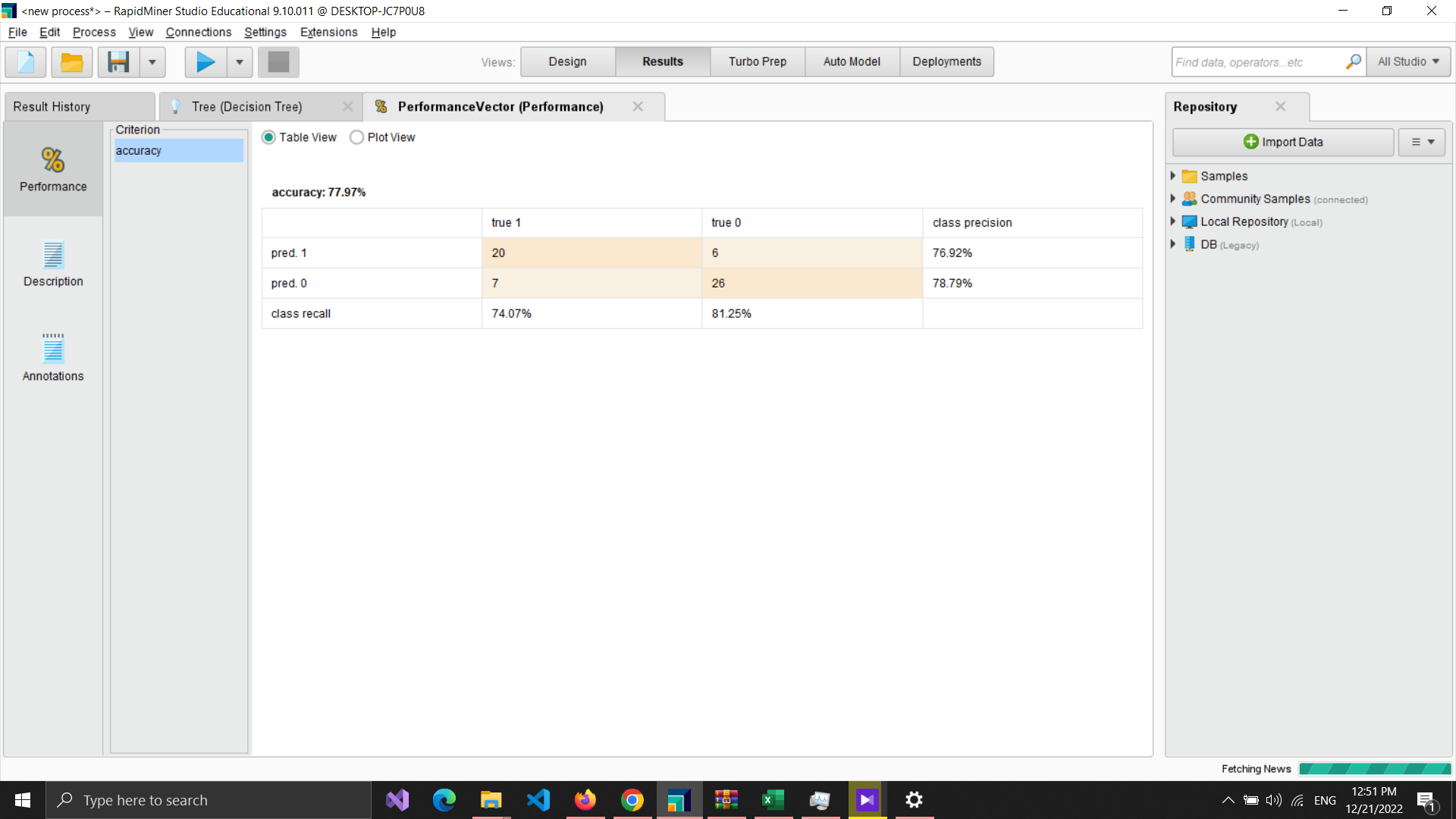
حال میاییم با استفاده از apply model درخت تصمیم را برحسب مدلی ک بهش میدهیم که اینجا برحسب ابزار split data داده ها رو به دو بخش تقسیم میکنم که یک بخش رو به مدل میدهیم ک بر اساس آن تصمیم بگیرد که داده های اصلی رو چگونه تقسیم کند و به آن درصد میدهیم در ابزار split data داده ها رو split میکنیم به دوبخش زیر:



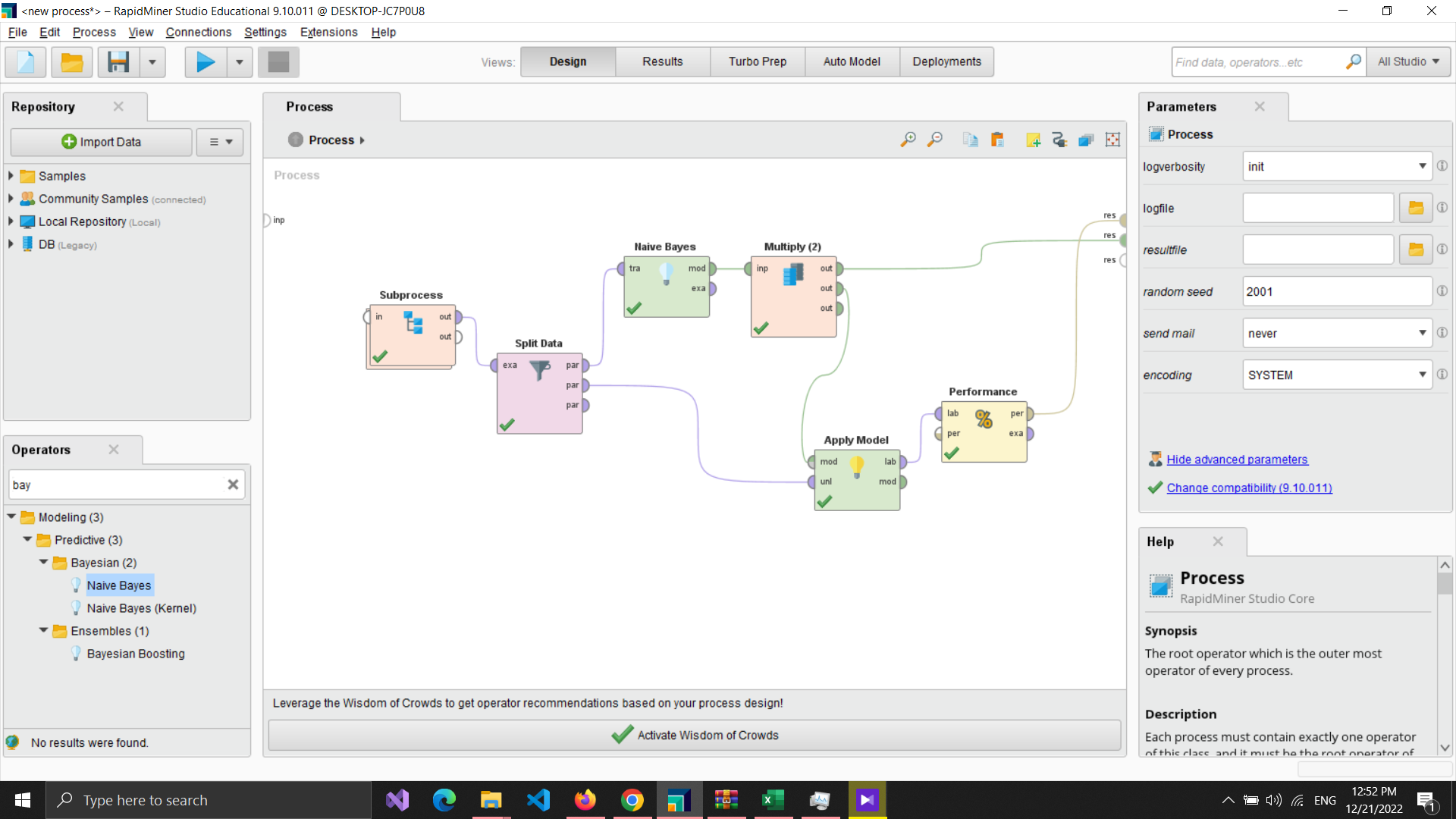
و به صورت زیر بهم وصلشان میکنیم.



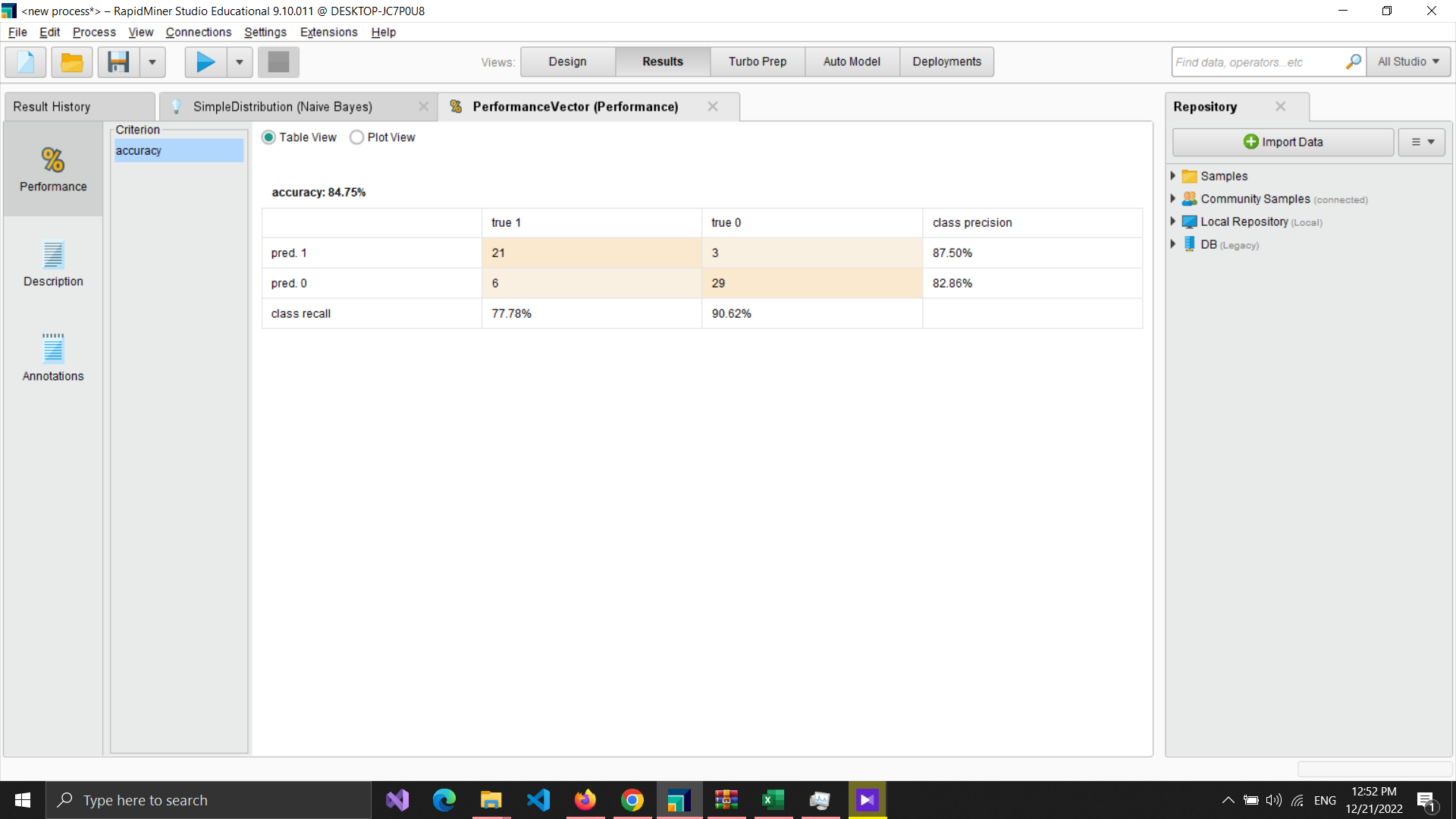
ابزار performance در آن ما می توانیم accuracy و recall مثلا این روش درخت تصمیم و روش بیزین رو ببنیم و باهم مقایسه کنیم که دردرخت تصمیم این داده ها به صورت زیر است.



و حال بجای ابزار درخت تصمیم ما از نوی بیزین استفاده میکنیم



و خروجی آن به صورت زیر است.



که در اینجا میبینم درروش بیزین نسبت به روش درخت تصمیم accuracy و recall آن افزایش پیدا کرده است.